Appendix A. Additional Results

Here we provide additional results from the simulation experiments and exemplar analysis in Sections 4 and 5 of the main paper.

A.1 Experiment 1 Coverage Plots

Figure 1 contains coverage plots for the standard test data set, Figure 2 contains coverage plots for the narrow test data, and Figure 3 contains coverage plots for the double peaked test data. These plots were produced using EFDM, SNCM1, GMD1, and GMD2 on the 500 full training data sets of size $n = 100$. With the standard function test data, Figure 1, we see how the different methods achieve validity (i.e., the correct coverage level) in different ways. For instance, EFDM is the only method that identifies outliers low-amplitude functions in the middle of the data as outliers. Interestingly, GMD1 identifies outliers to exist in phase, but the functions that are even more extreme in phase are labeled as inliers. GMD2 does not demonstrate this behavior at all.

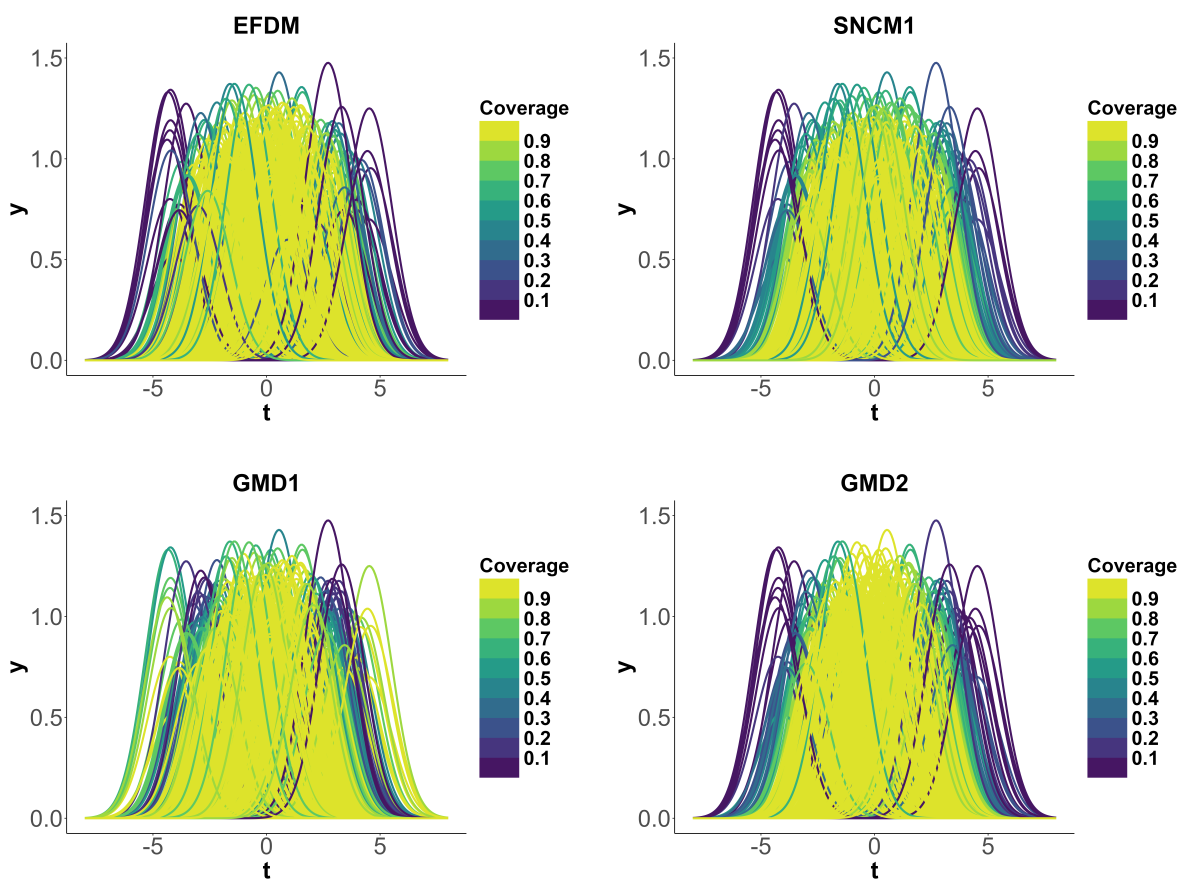

Figure 1: Plots of average coverage per test function in the standard test data for EFDM, SNCM1, GMD1, and GMD2 when $n = 100$.

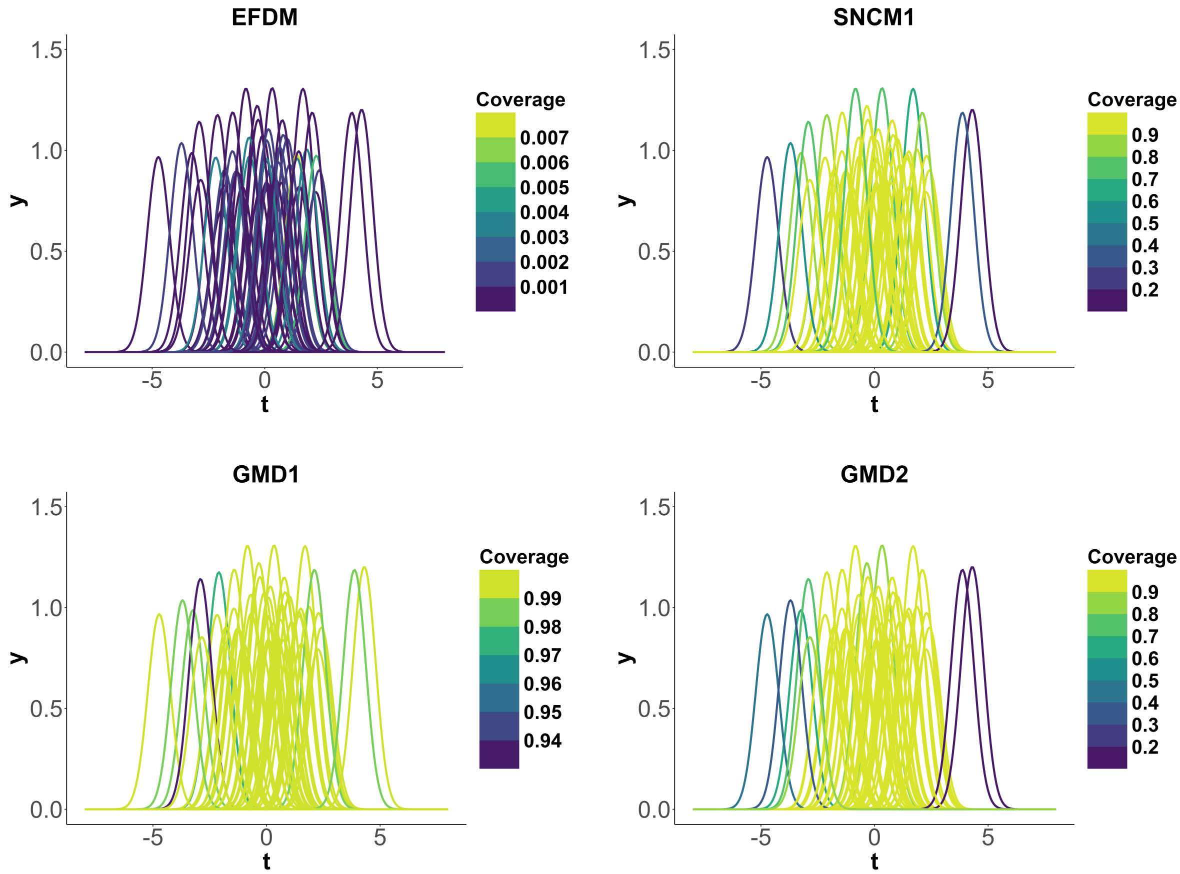

Figure 2: Plots of average coverage per test function in the narrow test data for EFDM, SNCM1, GMD1, and GMD2 when $n = 100$. Note the radically different color scales between the plots.

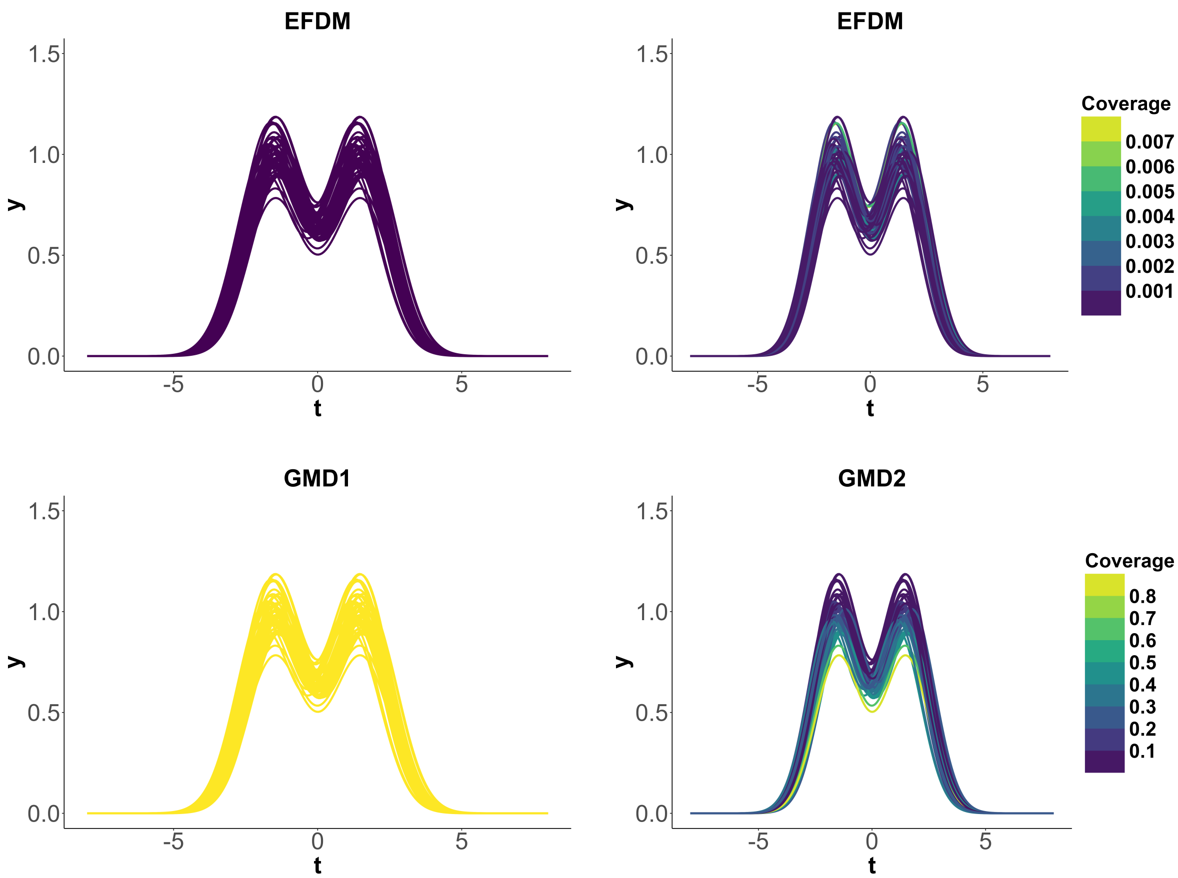

Figure 3: Plots of average coverage per test function in the double peaked test data for EFDM, SNCM1, GMD1, and GMD2 when $n = 100$. Color scales are not given for EFMD or GMD1 because all coverage values are identical. For EFDM, the coverage values are 0 while for GMD1, they are 1.

A.2 Experiment 2 Metrics

The tables in this section give the mean and standard deviation values for the true positive rate (TPR), true negative rate (TNR), positive predictive value (PPV), and negative predictive value (NPV) for experiment 2 at two different significance levels. Recall that outliers are considered the positive class and inliers the negative class when computing these metrics.

	5% Outliers in Test Data			10% Outliers in Test Data		
Method	$n = 50$	$n = 100$	$n = 250$	$n = 50$	$n = 100$	$n = 250$
EFDM	**0.854(0.16)**	**0.997(0.04)**	**1(0)**	**0.846(0.12)**	**0.997(0.03)**	**1(0)**
SNCM1	0.017(0.08)	0.011(0.05)	0.006(0.03)	0.017(0.07)	0.007(0.03)	0.006(0.03)
SNCM2	0.024(0.09)	0.015(0.06)	0.006(0.04)	0.022(0.08)	0.013(0.05)	0.007(0.03)
GMD1	0.0004(0.009)	0(0)	0(0)	0(0)	0(0)	0(0)
GMD2	0.376(0.35)	0.486(0.34)	0.575(0.30)	0.393(0.33)	0.487(0.32)	0.576(0.28)
GMD3	–	–	0.554(0.36)	–	–	0.546(0.34)
GMDM	–	–	0.624(0.34)	–	–	0.629(0.31)

Table 1: Mean(SD) TPR values for experiment 2 using $\alpha = 0.05$.

	5% Outliers in Test Data			10% Outliers in Test Data		
Method	$n = 50$	$n = 100$	$n = 250$	$n = 50$	$n = 100$	$n = 250$
EFDM	**0.999(0.01)**	**1(0)**	**1(0)**	**1(0)**	**1(0)**	**1(0)**
SNCM1	0.042(0.13)	0.032(0.09)	0.022(0.07)	0.04(0.11)	0.024(0.07)	0.02(0.05)
SNCM2	0.048(0.13)	0.037(0.1)	0.022(0.07)	0.049(0.11)	0.038(0.09)	0.019(0.05)
GMD1	0.0004(0.01)	0(0)	0(0)	0(0)	0(0)	0(0)
GMD2	0.629(0.35)	0.781(0.27)	0.866(0.19)	0.63(0.34)	0.782(0.25)	0.86(0.17)
GMD3	–	–	0.841(0.24)	–	–	0.839(0.21)
GMDM	–	–	0.896(0.18)	–	–	0.897(0.15)

Table 2: Mean(SD) TPR values for experiment 2 using $\alpha = 0.10$.

	5% Outliers in Test Data			10% Outliers in Test Data		
Method	$n = 50$	$n = 100$	$n = 250$	$n = 50$	$n = 100$	$n = 250$
EFDM	0.951(0.05)	0.95(0.04)	**0.952(0.03)**	0.949(0.05)	0.947(0.04)	0.948(0.03)
SNCM1	0.949(0.05)	**0.951(0.04)**	0.95(0.03)	0.948(0.05)	**0.95(0.04)**	0.948(0.03)
SNCM2	0.95(0.05)	0.95(0.04)	0.949(0.03)	0.949(0.05)	0.948(0.04)	0.947(0.03)
GMD1	0.949(0.05)	0.947(0.04)	0.949(0.03)	0.95(0.05)	0.948(0.04)	**0.949(0.03)**
GMD2	**0.952(0.05)**	0.949(0.04)	0.95(0.03)	**0.951(0.05)**	0.949(0.04)	**0.949(0.03)**
GMD3	–	–	**0.952(0.03)**	–	–	**0.949(0.03)**
GMDM	–	–	0.944(0.03)	–	–	0.943(0.03)

Table 3: Mean(SD) TNR values for experiment 2 using $\alpha = 0.05$.

	5% Outliers in Test Data			10% Outliers in Test Data		
Method	$n = 50$	$n = 100$	$n = 250$	$n = 50$	$n = 100$	$n = 250$
EFDM	0.899(0.07)	**0.899(0.06)**	**0.901(0.04)**	0.896(0.07)	**0.897(0.06)**	0.898(0.05)
SNCM1	0.900(0.07)	**0.899(0.06)**	0.900(0.04)	0.899(0.07)	**0.897(0.06)**	0.898(0.04)
SNCM2	**0.901(0.06)**	0.898(0.06)	0.900(0.04)	0.899(0.07)	0.896(0.06)	0.899(0.04)
GMD1	0.900(0.08)	0.896(0.06)	0.901(0.05)	**0.901(0.07)**	**0.897(0.06)**	0.900(0.05)
GMD2	0.901(0.07)	0.897(0.06)	**0.901(0.04)**	**0.901(0.07)**	0.895(0.06)	**0.901(0.04)**
GMD3	–	–	0.900(0.04)	–	–	0.899(0.05)
GMDM	–	–	0.881(0.04)	–	–	0.882(0.04)

Table 4: Mean(SD) TNR values for experiment 2 using $\alpha = 0.10$.

	5% Outliers in Test Data			10% Outliers in Test Data		
Method	$n = 50$	$n = 100$	$n = 250$	$n = 50$	$n = 100$	$n = 250$
EFDM	**0.588(0.26)**	**0.584(0.20)**	**0.567(0.16)**	**0.708(0.20)**	**0.716(0.16)**	**0.708(0.14)**
SNCM1	0.007(0.03)	0.007(0.03)	0.003(0.02)	0.013(0.05)	0.009(0.06)	0.008(0.04)
SNCM2	0.011(0.04)	0.01(0.04)	0.004(0.02)	0.021(0.08)	0.014(0.05)	0.01(0.04)
GMD1	0.0005(0.01)	0(0)	0(0)	0(0)	0(0)	0(0)
GMD2	0.282(0.29)	0.331(0.23)	0.385(0.20)	0.426(0.32)	0.5(0.24)	0.558(0.19)
GMD3	–	–	0.357(0.23)	–	–	0.496(0.24)
GMDM	–	–	0.363(0.19)	–	–	0.53(0.19)

Table 5: Mean(SD) PPV values for experiment 2 using $\alpha = 0.05$.

	5% Outliers in Test Data			10% Outliers in Test Data		
Method	$n = 50$	$n = 100$	$n = 250$	$n = 50$	$n = 100$	$n = 250$
EFDM	**0.417(0.19)**	**0.397(0.16)**	**0.375(0.11)**	**0.571(0.18)**	**0.556(0.15)**	**0.545(0.11)**
SNCM1	0.013(0.04)	0.012(0.04)	0.008(0.03)	0.025(0.06)	0.017(0.05)	0.017(0.04)
SNCM2	0.014(0.04)	0.015(0.04)	0.009(0.03)	0.031(0.07)	0.027(0.06)	0.016(0.04)
GMD1	0.0003(0.01)	0(0)	0(0)	0(0)	0(0)	0(0)
GMD2	0.278(0.20)	0.315(0.15)	0.333(0.10)	0.419(0.23)	0.476(0.14)	0.510(0.11)
GMD3	–	–	0.325(0.12)	–	–	0.493(0.12)
GMDM	–	–	0.297(0.08)	–	–	0.472(0.10)

Table 6: Mean(SD) PPV values for experiment 2 using $\alpha = 0.10$.

	5% Outliers in Test Data			10% Outliers in Test Data		
Method	$n = 50$	$n = 100$	$n = 250$	$n = 50$	$n = 100$	$n = 250$
EFDM	**0.992(0.01)**	**1(0)**	**1(0)**	**0.982(0.01)**	**1(0)**	**1(0)**
SNCM1	0.948(0.004)	0.948(0.003)	0.948(0.002)	0.897(0.01)	0.896(0.004)	0.896(0.004)
SNCM2	0.949(0.004)	0.948(0.003)	0.948(0.002)	0.897(0.01)	0.896(0.01)	0.896(0.004)
GMD1	0.947(0.003)	0.947(0.003)	0.947(0.002)	0.895(0.01)	0.895(0.005)	0.895(0.003)
GMD2	0.967(0.02)	0.973(0.02)	0.977(0.02)	0.936(0.03)	0.945(0.03)	0.954(0.03)
GMD3	–	–	0.976(0.02)	–	–	0.951(0.04)
GMD4	–	–	0.98(0.02)	–	–	0.96(0.03)

Table 7: Mean(SD) NPV values for experiment 2 using $\alpha = 0.05$.

	5% Outliers in Test Data			10% Outliers in Test Data		
Method	$n = 50$	$n = 100$	$n = 250$	$n = 50$	$n = 100$	$n = 250$
EFDM	**1(0)**	**1(0)**	**1(0)**	**1(0)**	**1(0)**	**1(0)**
SNCM1	0.947(0.01)	0.946(0.01)	0.946(0)	0.894(0.01)	0.892(0.01)	0.892(0.01)
SNCM2	0.947(0.01)	0.946(0.01)	0.946(0)	0.895(0.01)	0.893(0.01)	0.892(0.01)
GMD1	0.944(0.01)	0.944(0.004)	0.945(0.003)	0.890(0.01)	0.889(0.01)	0.890(0.01)
GMD2	0.98(0.02)	0.988(0.01)	0.993(0.01)	0.959(0.04)	0.975(0.03)	0.984(0.02)
GMD3	–	–	0.991(0.01)	–	–	0.981(0.02)
GMDM	–	–	0.994(0.01)	–	–	0.988(0.02)

Table 8: Mean(SD) NPV values for experiment 2 using $\alpha = 0.10$.

A.3 Additional Exemplar Analysis Results

Figure 4 displays the results of the exemplar analysis presented in the main paper. In the figure, functions are colored by their p-values so that darker functions are more likely to be labeled as outliers.

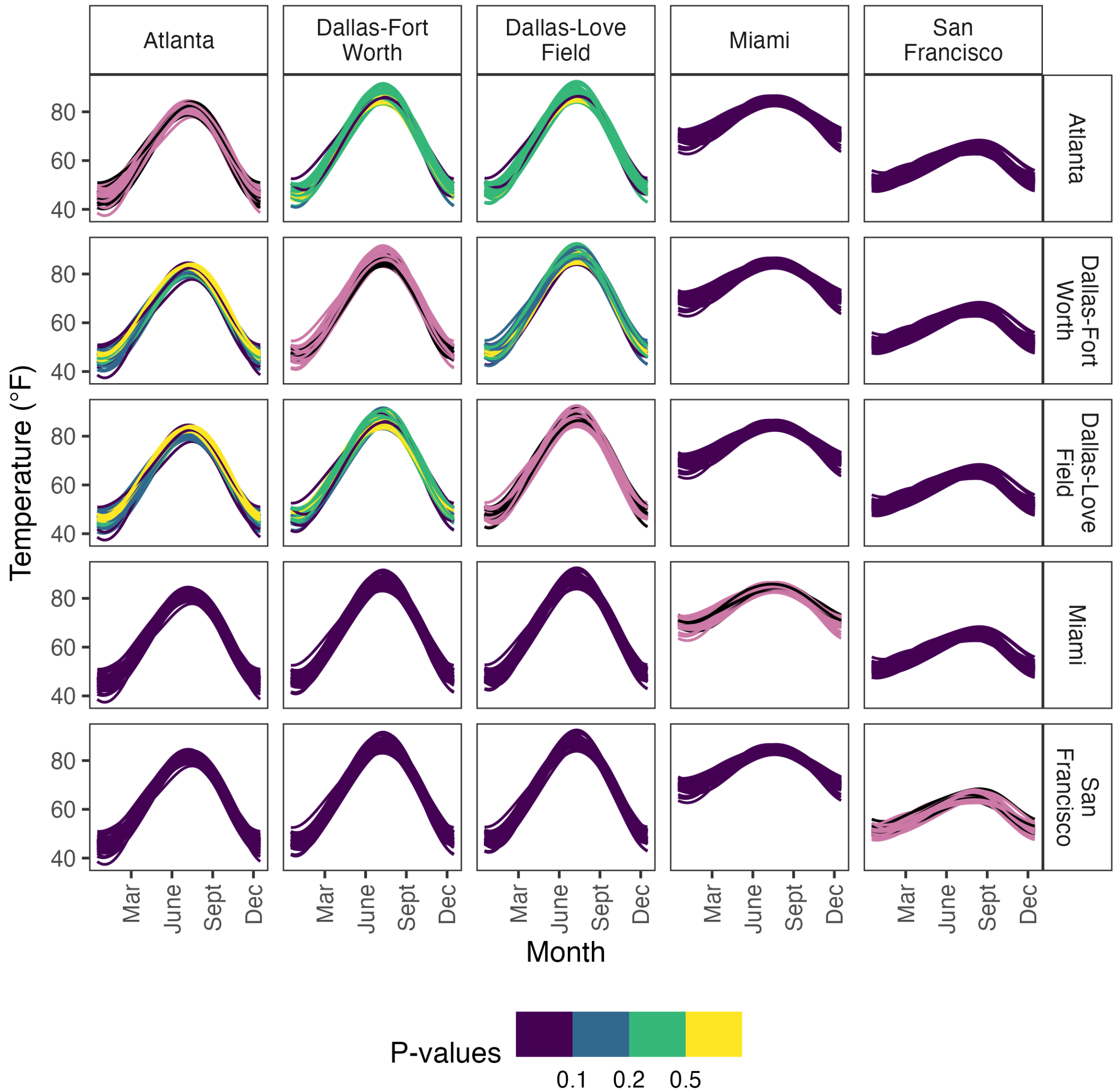

Figure 4: Each row of the figure corresponds to using the site listed on the right as the training and calibration data. Each column corresponds to the site listed on top as the test site. When the training site and test site are the same (i.e. the plots along the diagonal) the plot consists of just the training and calibration data. The pink functions are the training functions and the black functions are the calibration functions. When the training site and the test site are different the plot shows the test data. Each function is color coded by the p-value bin it falls in. Note that bins used to differentiate the color of the functions are not equally sized.